

AUDIO PROCESSING SYSTEM AND METHOD FOR CLASSIFYING SPEAKERS IN AUDIO DATA

5

ABSTRACT OF THE DISCLOSURE

An audio processing system and method for classifying speakers in audio data using a discriminatively-trained classifier. In general, the audio processing system inputs audio data containing unknown speakers and outputs frame tags whereby each tag represents an individual speaker. The audio processing system includes a training system for training a discriminatively-trained classifier (such as a time-delay neural network) and a speaker classification system for using the classifier to segment and classify the speakers. The audio processing method includes two phases. A training phase discriminatively trains the classifier on a speaker training set containing known speakers and produces fixed classifier data. A use phase uses the fixed classifier data in the discriminatively-trained classifier to produce anchor model outputs for every frame of speech in the audio data. The anchor model outputs are mapped to frame tags so that all speech corresponding to a single frame tag comes from a single speaker.